Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2016 April ; 25(4): 613–623. doi:10.1158/1055-9965.EPI-15-0225.

What Predicts an Advanced-Stage Diagnosis of Breast Cancer? Sorting Out the Influence of Method of Detection, Access to Care, and Biologic Factors

Joseph Lipscomb[1], **Steven T. Fleming**[2], **Amy Trentham-Dietz**[3], **Gretchen Kimmick**[4], **Xiao-Cheng Wu**[5], **Cyllene R. Morris**[6], **Kun Zhang**[7], **Robert A. Smith**[8], **Roger T. Anderson**[9], **Susan A. Sabatino**[10] **on behalf of the Centers for Disease Control and Prevention National Program of Cancer Registries Patterns of Care Study Group**

[1]Department of Health Policy and Management, Rollins School of Public Health, Winship Cancer Institute, Emory University, Atlanta, Georgia.

[2]Department of Epidemiology, University of Kentucky College of Public Health, Lexington, Kentucky.

[3]University of Wisconsin Carbone Cancer Center, Madison, Wisconsin.

[4]Department of Internal Medicine, Medical Oncology, Duke University Medical Center and Multidisciplinary Breast Cancer Program, Duke Cancer Institute, Durham, North Carolina.

[5]Epidemiology Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, Louisiana.

[6]California Cancer Registry, Institute for Population Health Improvement, UC Davis Health System, Sacramento, California.

[7]Department of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, Georgia.

[8]American Cancer Society, Atlanta, Georgia.

Corresponding Author: Joseph Lipscomb, Emory University, 1518 Clifton Road, Rm 720, Atlanta, GA 30322. Phone: 404-727-4513; Fax: 404-727-9198; jlipsco@emory.edu.

Authors' Contributions

Conception and design: J. Lipscomb, S.T. Fleming, A. Trentham-Dietz, G. Kimmick, X.-C. Wu, R.T. Anderson

Development of methodology: J. Lipscomb, A. Trentham-Dietz, X.-C. Wu, K. Zhang

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Lipscomb, A. Trentham-Dietz, X.-C. Wu

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Lipscomb, S.T. Fleming, G. Kimmick, X.-C. Wu, C.R. Morris, K. Zhang, R.A. Smith

Writing, review, and/or revision of the manuscript: J. Lipscomb, S.T. Fleming, A. Trentham-Dietz, G. Kimmick, X.-C. Wu, C.R. Morris, R.A. Smith, R.T. Anderson, S.A. Sabatino

Study supervision: J. Lipscomb, X.-C. Wu

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

[9]Department of Public Health Sciences, University of Virginia School of Medicine, and UVA Cancer Center, Charlottesville, Virginia.

[10]Division of Cancer Prevention and Control, U.S. Centers for Disease Control and Prevention, Atlanta, Georgia.

Abstract

Background: Multiple studies have yielded important findings regarding the determinants of an advanced-stage diagnosis of breast cancer. We seek to advance this line of inquiry through a broadened conceptual framework and accompanying statistical modeling strategy that recognize the dual importance of access-to-care and biologic factors on stage.

Methods: The Centers for Disease Control and Prevention–sponsored Breast and Prostate Cancer Data Quality and Patterns of Care Study yielded a seven-state, cancer registry–derived population-based sample of 9,142 women diagnosed with a first primary *in situ* or invasive breast cancer in 2004. The likelihood of advanced-stage cancer (American Joint Committee on Cancer IIIB, IIIC, or IV) was investigated through multivariable regression modeling, with base-case analyses using the method of instrumental variables (IV) to detect and correct for possible selection bias. The robustness of base-case findings was examined through extensive sensitivity analyses.

Results: Advanced-stage disease was negatively associated with detection by mammography ($P < 0.001$) and with age < 50 ($P < 0.001$), and positively related to black race ($P = 0.07$), not being privately insured [Medicaid ($P = 0.01$), Medicare ($P = 0.04$), uninsured ($P = 0.07$)], being single ($P = 0.06$), body mass index > 40 ($P = 0.001$), a HER2 type tumor ($P < 0.001$), and tumor grade not well differentiated ($P < 0.001$). This IV model detected and adjusted for significant selection effects associated with method of detection ($P = 0.02$). Sensitivity analyses generally supported these base-case results.

Conclusions: Through our comprehensive modeling strategy and sensitivity analyses, we provide new estimates of the magnitude and robustness of the determinants of advanced-stage breast cancer.

Impact: Statistical approaches frequently used to address observational data biases in treatment-outcome studies can be applied similarly in analyses of the determinants of stage at diagnosis.

Introduction

An advanced-stage diagnosis of breast cancer has long been associated with significantly poorer survival outcomes (1). Recent data show that women diagnosed at American Joint Committee on Cancer (AJCC) stage I have an overall 5-year relative survival rate near 100%, whereas the rate for those diagnosed at stage IV is 24% (2). Over two decades of investigations into the determinants of an advanced-stage diagnosis have yielded important findings.

Screening mammography has been consistently associated with earlier-stage detection of breast cancer, both in clinical trials (3–7) and in day-to-day practice (8–10). This is notwithstanding important complicating factors, including disagreement about appropriate

screening strategies (11, 12) and variability in mammographic test sensitivity driven by certain biologic factors including mammographic breast density (13–15).

There are race/ethnicity differences in breast cancer stage at diagnosis (16–30), with African American women significantly more likely than white women to have a late-stage diagnosis. Insurance status is a significant independent predictor of stage, with women who are uninsured or enrolled in Medicaid less likely to access screening mammography (31, 32) and more likely to be diagnosed at later stage (20, 21, 33–35). There is a complex interplay involving the presence of comorbidities, access to care, detection by mammography, and stage (36, 37).

Taken together, these studies have significantly contributed to our understanding of factors associated with an advanced stage diagnosis of breast cancer. However, there are certain methodological considerations, not explored to date, with potentially important implications for the specification of models and interpretation of findings.

First, in virtually all studies, the likelihood of an advanced-stage diagnosis has been analyzed through a single-equation (typically logistic) regression model in which explanatory variables, including method of detection, were all regarded as independent, exogenous predictors of stage. In reality, the detection method is not a fixed, predetermined variable in the same sense as the individual's age or race/ethnicity. Rather, it can be regarded, as indeed it has been in the screening mammography trials, as an "exposure" influencing the "outcome" of stage at diagnosis.

Second, many analyses have relied heavily on cancer registry sources that do not routinely include several potentially important predictors of both stage and method of detection. Such unobserved variables may include breast density, which influences both mammography sensitivity (14, 15) and tumor aggressiveness (38); whether the woman is taking hormone replacement therapy, which may affect tumor development (39) and the perceived importance of regular screening; whether the woman has a family history of breast cancer; the nature of the woman's health care system (e.g., managed care vs. fee-for-service), which may influence both screening rates and the effectiveness of follow-up care (40); and certain health behaviors, e.g., excess alcohol consumption (41).

Third, to the extent such unobserved variables are important predictors of both stage and detection method, the statistical problem of endogeneity arises: the error structures for the regression models predicting stage and predicting method are then correlated (because they contain common unobserved variables). Without corrections for such potential endogeneity, estimates of the impact of predictors—such as method, race/ethnicity, and insurance—on stage are subject to bias (42).

In this article, we bring a new conceptual framework and accompanying statistical modeling strategy—built primarily around the method of instrumental variables (IV; refs. 42–47)—to a much-analyzed question: What predicts an advanced-stage diagnosis of breast cancer? Of particular interest is whether findings to date regarding the impact of method of detection, race/ethnicity, and insurance status on stage are sustained within this expanded framework.

Materials and Methods

Conceptual framework

Our maintained hypothesis regarding the causal factors leading to the breast cancer stage recorded at diagnosis (stage) is depicted in Fig. 1. Method of detection (method) plays a central role, and we specify two sets of variables that may influence both stage and method. One set contains variables associated with the woman's access to and utilization of health care (e.g., insurance status). The second set consists of variables associated with the aggressiveness and speed of the tumor's biologic development (e.g., grade), the sensitivity of detection methods (e.g., histology), or both [e.g., body mass index (BMI)]. Each set includes variables that, depending on the available data sources, may be observable (e.g., marital status) or unobservable (e.g., breast density, menopausal hormone therapy) to the investigator.

Empirical basis

Data sources.—The principal source of data is the Breast and Prostate Cancer Data Quality and Patterns of Care Study (POC-BP), funded by the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC) and involving investigators affiliated with population-based registries in seven states (California, Georgia, Kentucky, Louisiana, North Carolina, Minnesota, and Wisconsin) and the CDC. Institutional Review Board approval was obtained from all participating states, academic institutions, and government agencies.

Over 2007 to 2009, the POC-BP sampled NPCR patients diagnosed in 2004, with intensive re-abstraction of medical records from hospitals and outpatient facilities (including pathology laboratories, radiation facilities, surgical centers, and physician offices).

Patient eligibility and selection.—Our analyses included women ≥ 20 years of age diagnosed in 2004 with microscopically confirmed *in situ* or invasive primary breast cancer (International Classification of Disease-Oncology, 3rd Edition, site codes C50.0-C50.9) with no previous cancer diagnosis and meeting other standard exclusion criteria. Cases diagnosed at Veterans Affairs hospitals were excluded because of data availability limitations.

Cases were selected from the NPCR registries through single-stage random sampling stratified by race/ethnicity in all states and by other factors that varied by state (e.g., by urban/rural status in Georgia). A detailed account of data collection and quality assessment for POC-BP has been reported (48).

Derivation of variables

Stage at diagnosis.—Patients were assigned an AJCC (Sixth Edition TNM) stage based on the collaborative stage algorithm in effect for 2004 diagnoses. There is wide variability in how previous studies have defined advanced (or late) stage of breast cancer: III or IV (24, 27, 30); II, III, or IV (19, 21, 22, 29); IIB, III, or IV (36). In response, we defined advanced stage on the basis of the pattern of decline by AJCC stage in 5-year overall survival rates. SEER*Stat analyses (2) on cases diagnosed in 2004–2010 and followed through 2011

yielded these 5-year survival percentages by AJCC stage: 0 (95.4), I (91.9), IIA (86.8), IIB (81.9), IIIA (76.9), IIIB (54.0), IIIC (59.6), and IV (20.8). Given the sharp drop-off between IIIA and IIIB, we designated "advanced" stage as diagnosis at IIIB, IIIC, or IV. All others were diagnosed at an "earlier" stage.

Method of detection.—We defined a two-level variable, mammography and other, where "other" included detection by clinical breast examination (CBE), breast self-exam (BSE), or signs/symptoms. Each patient was assigned a detection method based on a detailed review of medical records at the site(s) where she received breast cancer care. The intent was to capture the initial detection-related event that triggered steps toward a definitive diagnosis. Thus, if an initial BSE led to a mammogram, which led eventually to a breast cancer diagnosis, the coded method of detection would be BSE; see Table 1 for more detail.

Factors associated with health care access and utilization.—These included race/ethnicity, insurance status, comorbidity status [based on Piccirillo's Adult Comorbidity Evaluation (ACE) instrument (49)], marital status, age, and several area-level variables constructed from 2000 US Census data: urban/rural status, poverty status, and education status. These categorical variables are operationally defined in Table 1.

Biologic factors associated with tumor progression and detection.—In addition to race/ethnicity and age, we included BMI, tumor grade, and a constructed variable "mole-subtype" based on the patient's combined estrogen receptor (ER), progesterone receptor (PR), and HER2 status and intended to approximate the molecular subtype of the breast tumor (50). Specifically, a patient here may be "Luminal A" (ER^+ and/or PR^+, $HER2^-$), "Luminal B" (ER^+ and/PR^+, $HER2^+$), "triple negative" (ER^-, PR^-, $HER2^-$), or "HER2 Type" (ER^-, PR^-, $HER2^+$).

Statistical analyses

Predicting stage.—Our overall strategy is to compare conventional single-equation regression models with formulations designed to detect and correct for selection bias, under a variety of assumptions. Prototypically, the single-equation models will be binary logistic regressions with the log-odds of an advanced-stage diagnosis being a function of method of detection plus some combination of patient-level access/utilization factors and biologic factors. From the standpoint of Fig. 1, such single-equation models correspond to a conceptual framework that omits both arrows directed at method of detection.

To test and correct for any selection bias, our primary approach is the method of instrumental variables using the two-stage residual inclusion (2SRI) model (45–47, 51), which is especially well suited for nonlinear estimation, as here. Figure 2 is a transformation of Fig. 1 that depicts key aspects of our 2SRI econometric model, including the main observable and unobservable variables thought to be at play. To execute the 2SRI model, one estimates a first-stage regression in which the likelihood the patient receives the "intervention" (here, mammography) is a function of all observable predictors posited to influence the patient's "outcome" (here, advanced vs. earlier stage) plus instrumental variable(s), hypothesized to influence the patient's selection into intervention but not her

outcome (except through their impact on choice of intervention). In the second-stage regression, the likelihood of advanced stage becomes a function of method of detection, the hypothesized access/utilization factors and biologic factors, and a variable consisting of the residuals computed from the first-stage regression and intended to both indicate the degree of selection bias and correct for it (42, 46).

Because the residual is a computed variable, thereby reflecting sampling error, we employed bootstrapping (100 iterations) to upwardly adjust the standard errors of coefficient estimates in the second-stage regression.

The IVs deployed here are seen in Fig. 2 and discussed further in Table 2: the patient's state of residence [mammography rates vary across the seven states, reflecting an underlying geographic variability in screening uptake (52)]; histology [because mammography is less sensitive for lobular tumors (52), while histology itself is posited not to be an important predictor of tumor aggressiveness after adjusting for other biologic variables]; and a constructed variable, mammography-capacity, indexing a woman's physical access to mammography in her county of residence. Because of state-imposed confidentiality requirements, mammography-capacity could not be computed for Minnesota.

The strength of the IVs is indexed by the magnitude of the F-statistic for the null they are jointly 0 in the first-stage regression; a frequently used, if informal, benchmark is that $F \geq 10$ (42, 53).

Additional statistical considerations.—For three predictor variables with substantial missing observations among the cases potentially available for analysis—mole-subtype (29.8%), BMI (22.0%), and grade (8.6%)—we used multiple imputation (MI) to assign values (54, 55). (No other variable was missing more than 3%, and most were missing under 1%.) When MI was applied to the 2SRI models, standard errors were constructed to reflect the sampling variability arising from both the computed residuals and the imputation process; see Supplementary Materials (Section A).

In regressions using data from single-stage sampling, where the weights are a function of predictor variables (here, for example, race–ethnicity), using sample weights can, at a minimum, reduce statistical precision (56); consequently, we did not weight the data in the base-case.

With binary logistic regression used throughout, our complementary measures of model performance are the coefficient of concordance (c-statistic) and the Hosmer–Lemeshow (H-L) goodness-of-fit test statistic (57). H-L indexes how well predicted and observed event rates (here advanced-stage) match up in subgroups (typically deciles) of the sample; the closer the match, the higher the *P* value.

Regression results are expressed as adjusted ORs with corresponding 95% confidence intervals; *P* values are two-sided, with $P \leq 0.05$ as a benchmark for appraising statistical importance. Analyses used Stata, version 13.0 (Stata Corporation), and SAS, version 9.2 (SAS Institute, Cary, NC).

Base-case model specification.—In summary, our base-case regression model for stage at diagnosis was a 2SRI specification estimated (1) without sample weights; (2) with missing values imputed for BMI, grade, and mole-subtype (but not other predictors); and (3) with the IV mammography-capacity included. As detailed below, we conducted extensive sensitivity analyses, including propensity score weighting as an alternative to IV for selection bias reduction (58–62).

Results

Descriptive and bivariate analyses

From a total 11,643 qualifying breast cancer cases, re-abstraction was successfully completed on 9,142. Among these, 212 were not assigned an AJCC stage, and 230 additional patients were missing information on method of detection. Hence, 8,700 cases were potentially available for analysis (because, following standard practice, we did not apply MI to our two dependent variables, stage and method). Under base-case modeling assumptions (which exclude Minnesota), the corresponding sample has 7,503 patients. Among these, 762 (10.2%) were diagnosed at advanced stage (Table 1), and 3,718 (49.6%) cases overall were detected by mammography (Table 2).

For most predictor variables in Table 1, there were notable differences in the distribution of patients between advanced versus earlier stage at each level of the variable, and the corresponding unadjusted ORs from the binary logistic regression of the variable on stage were significant at $P < 0.05$ in most cases. For example, among those detected by mammography, only 2.6% were at advanced stage, whereas for those detected by some other method, 17.6% were advanced stage; the unadjusted OR for detection by mammography (vs. other method) being associated with advanced stage was 0.13 ($P < 0.001$).

Table 2 presents a parallel summary of information for the IVs. Of prime interest is the association between each IV and the likelihood of detection by mammography. For histology and mammography-capacity, the unadjusted ORs were significant, in the expected directions; although the state variable was not as strongly associated with method, we elected to retain it as an IV, given *a priori* expectations about geographic variations in screening practices.

Inferences from first-stage regression in 2SRI model

The motivating purpose of the first-stage regression is to derive the “Method of Detection Bias Correction Factor” (Fig. 2)—that is, a variable consisting of that model’s residuals which then enters the second-stage regression for stage. Regarding the statistical strength of the IVs, the F statistic (8 df) for the null that state, histology, and mammography-capacity are jointly noninfluential was 105.7 ($P < 0.001$), well above the benchmark of $F \geq 10$. This estimated first-stage model is discussed in Supplementary Materials (Section B).

Base-case model for determinants of advanced-stage disease

The right-hand portion of Table 3 displays the estimated second-stage regression for the base-case 2SRI model. The likelihood of an advanced-stage diagnosis is strongly negatively

related to detection by mammography (OR = 0.04; $P < 0.001$). Of note, the first-stage residual is significantly positive (OR = 3.89; $P < 0.02$), consistent with selection bias in the "allocation" of women to mammography versus other.

Advanced stage was positively associated with being black (OR = 1.16; $P = 0.07$); not being privately insured, with Medicaid (OR = 1.48; $P = 0.01$) and Medicare (OR = 1.29; $P = 0.04$), and being uninsured (OR = 1.60; $P = 0.07$); being single (OR = 1.28; $P = 0.06$); having BMI ≥ 40 (OR = 1.62; $P = 0.001$); having a tumor of HER2 Type (OR = 1.40; $P < 0.001$); and having tumor grade that is moderately differentiated (OR = 2.36; $P < 0.001$) or else poorly or undifferentiated (OR = 3.91; $P < 0.001$). Advanced stage was negatively related to being diagnosed at age < 40 (OR = 0.48; $P = 0.001$) or between ages 40 and 49 (OR = 0.57; $P < 0.001$).

Results from the corresponding single-equation multivariable regression are in the left-hand portion of Table 3. Notwithstanding the significant bias correction term in the 2SRI model, there was general concordance in findings from the two models. The models had comparable within-sample predictive ability (c = 0.797 and 0.796), though the 2SRI model had a notably better HL statistic ($P = 0.88$ vs. 0.53).

Sensitivity analyses around the base-case

Multiple model variants were analyzed where, in each case, we altered one key base-case provision while retaining the others; see Supplementary Tables S1–S6 in Supplementary Materials (Section D). These variants included models that (i) excluded the biologic variables associated with tumor progression (Supplementary Table S1); (ii) did not impute missing values (Supplementary Table S2); (iii) did employ the sample weights (Supplementary Table S3); (iv) excluded the IV mammography-capacity and thus included the Minnesota observations (Supplementary Table S4); and (v) used propensity score weighting as an alternative bias-reduction technique (Supplementary Table S5). For further appraisal of these estimated propensity score models, see Supplementary Table S6 and Supplementary Materials (Section C).

Overall, the findings from Supplementary Tables S1–S5 are broadly in tune with Table 3, but there are some notable differences, as discussed below.

Discussion

Guided by a new conceptual framework and statistical modeling strategy, this article re-examines a much-investigated question: What predicts an advanced-stage diagnosis of breast cancer?

We found that detection by mammography is significantly negatively related to an advanced-stage diagnosis. Across all multivariable single-equation and propensity score–adjusted models, the OR (mammography vs. other) for advanced stage ranged narrowly from 0.13 to 0.15. Across the IV models, there was a consistent pattern: this adjusted OR was in the (much lower) 0.03–0.05 range, whereas the OR for the residual predictor variable—indicating the magnitude of selection bias—was in the 2.73–6.92 range. An OR > 1 implies

a positive relationship between the algebraic sign of the residuals and the likelihood of advanced stage. We posit this reflects the fact (Table 1) that over half of all tumors were detected by some other method, thus generating negative residuals from the first-stage regression, and that over 82% of these were earlier stage. In effect, the estimated 2SRI model appropriately down-weighted the credit given to other methods for "detecting" earlier stage tumors.

In interpreting these results, it is noteworthy that relatively few studies have examined the impact of method of detection on stage at diagnosis (as opposed to the more common scenario of examining the relationship between observed or reported screening behavior on stage). Analyzing data from three screening trials, Shen and colleagues (63) found a "clear shift toward earlier stage" in cancers detected by mammography. This is in line with results reported by Malmgren and colleagues (9) from a prospective cohort study of women aged 40 to 49 diagnosed across 1990–2008, from a Wisconsin study of cases diagnosed across 1987–1990 (64), and from a 2001–2003 study of cases diagnosed in Detroit and Los Angeles (23).

Because we cannot observe in the POC-BP data the actual frequency and timing of a woman's screening mammography (or her CBE or BSE), the method of detection variable is not a direct measure of the effectiveness of mammography (or CBE or BSE) in averting an advanced-stage breast cancer. Thus, we cannot know for sure that a tumor diagnosed at advanced stage by other methods would have been found at an earlier stage if the woman had been getting regular mammograms. It is possible that she had been receiving mammography (at some rate), and a small but aggressive tumor was missed and subsequently emerged as an "interval" cancer of advanced stage. One role of the method of detection variable here is to account for the net influence of these unobserved (in our data) screening behaviors on stage at diagnosis. Further discussion about the role of the method of detection variable in these analyses, and its interpretation, can be found in Supplementary Materials (Section E).

Race–ethnicity

The associations of race/ethnicity with stage within 12 alternative predictive models (including the base-case and Supplementary Tables S1–S5) are explored further in Table 4, with several implications. First, while Hispanics and Asian/Pacific Islanders (API), but not blacks, were significantly less likely to be detected by mammography than whites (Supplementary Materials, Section B), the only significant race/ethnicity difference in stage was between black and white women. Second, the odds of blacks being diagnosed at a later stage than whites varied across models 1 to 12 in the following general way: the richer the set of included covariates, the less influential was the race/ethnicity variable.

This pattern of findings underscores that the estimated magnitude of a race/ethnicity effect depends on the overall maintained hypothesis embodied in the chosen statistical model for stage. That said, a significant black–white difference in breast cancer stage at diagnosis has been reported almost without exception in US studies to date (16–29). Some studies have found that black–white differences are significantly reduced after accounting for screening history (22–24), while others have not (25, 29). We conclude that a black–white difference

in stage is a robust finding, but one whose magnitude and interpretation can vary across data sources, study designs, time frames, and geographic settings.

Insurance

Across the models in Table 4, the trend is clear: women without private insurance were significantly more likely to be diagnosed at advanced stage. Being uninsured had the greatest adverse impact, followed by having Medicaid, and then Medicare.

These findings generally align with earlier estimates (20, 21, 33–35). However, among the privately insured, we could not distinguish fee-for-service and managed care, and there may be a differential impact of coverage regime on method of detection and stage (65, 66). Overall, the likely route by which insurance influences stage is to increase the likelihood of detection by mammography (32); in parallel, insurance may increase the odds of timely diagnosis and treatment following a positive screen (8, 40, 67). Not directly accounted for here is whether the woman had a regular source of health care or received a provider recommendation for screening (8, 40).

Comorbidity

Patients with severe comorbidity were much less likely to be detected by mammography (Supplementary Materials, Section B) and significantly more likely to be diagnosed at advanced stage in our single-equation models; however, while OR > 1 for severe comorbidity in all 2SRI models, it was generally not significant. As Fleming and colleagues (36) note, the consolidation of multiple comorbid conditions into a single metric, such as the ACE-27, may hide antagonistic effects of individual comorbidities on either screening or advanced stage (36). Yasmeen and colleagues (37) found that comorbidities were positively associated with mammography use and also with an advanced-stage diagnosis among women who were screened most frequently.

Socioeconomic factors

While lower SES has been associated with late-stage breast cancer diagnosis in several studies (68–70), the only significant effect here was that single women were generally at higher risk to advanced stage than married women. The area-level variables indexing education, poverty, and urban-rural status were not significant in any model variant.

Biologic factors

Variables hypothesized to be associated with the aggressiveness and speed of tumor development generally performed as expected. Across models (see Table 3 and Supplementary Tables S1–S5), advanced stage was positively related to whether the tumor was HER2 Type, the tumor grade was not well differentiated, and the woman was morbidly obese (BMI ≥ 40); advanced stage was negatively associated with a diagnosis under age 50. While women with triple-negative disease were significantly less likely to be detected by mammography (Supplementary Materials, Section B), triple-negative status was not independently associated with an advanced-stage diagnosis in any model. As indicated in Fig. 2, a potentially important variable here not available in the POC-BP data set was breast density.

Although the joint impact of these factors on stage-at-diagnosis has not been previously evaluated, earlier studies support portions of our findings. For example, Kerlikowske and colleagues (71) found that screen-detected cancers were higher among overweight/obese women, whereas rates of advanced-stage diagnosis increased across BMI groups, controlling for mammography use; however, an earlier study did not find a connection between BMI and stage for screen-detected cancers (64). The complex interplay involving hormonal status, postmenopausal hormone use, age, menopause, BMI, mammography use and sensitivity, and stage remains a ripe topic for investigation (15, 39, 71–77).

Concluding observations

In recent years, observational studies have examined the impact of various factors on breast cancer stage at diagnosis: method of detection (9); race, ethnicity, and socioeconomic variables (23); and biomedical variables such as BMI (72). This article adopts the perspective that the most conceptually and statistically defensible approach to understanding the influence of each such factor is to study them in concert (Figs. 1 and 2).

Overall, our findings about the determinants of advanced-stage align with those reported variously over the past two decades. What this study does provide, through its comprehensive modeling strategy and multiple sensitivity analyses, are new—and we think better grounded—estimates of the magnitude and statistical robustness of these posited predictors for stage.

What is needed going forward are continuing efforts to expand the empirical base so that the influence of potentially important unobservables (e.g., behavioral risk factors, health system effects, clinical variables not routinely collected in population-based studies) can be gauged in the context of an ever-more-richly specified model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are indebted to the editors and two anonymous reviewers for multiple constructive comments and recommendations that have sharpened the interpretation of study findings and their implications. They are also grateful to Lyn Almon, Judy Andrews, and Kevin Ward at the Georgia Center for Cancer Statistics at Emory University for analytical file development and interpretation; to Trevor Thompson in the CDC Division of Cancer Prevention and Control for database development; and to Jennifer Wike at the DB Consulting Group (a contractor to the CDC) for ongoing logistical and management support of this CDC-supported study.

Grant Support

From the Breast and Prostate Cancer Data Quality and Patterns of Care Study, supported by the CDC through cooperative agreements with the California Cancer Registry (Public Health Institute; 1-U01-DP000260; to R. Cress), Emory University (1-U01-DP000258; principal investigator, J. Lipscomb), Louisiana State University Health Sciences Center (1-U01-DP000253; principal investigator, X.-C. Wu), Minnesota Cancer Surveillance System (1-U01-DP000259; principal investigator, C.I. Perkins), Medical College of Wisconsin (1-U01-DP000261; principal investigator, J.F. Wilson), University of Kentucky (1-U01-DP000251; principal investigator, S.T. Fleming), and Wake Forest University (1-U01-DP000264; principal investigator, R.T. Anderson). J. Lipscomb was supported also by Cancer Center Support Grant P30CA138292 to the Winship Cancer Institute of Emory University from the NCI.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

References

1. Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, et al. (eds). SEER Cancer Statistics Review, 1975–2011. Bethesda, MD: NCI. Available from: http://seer.cancer.gov/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER web site, April 2014.
2. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data+Hurricane Katrina Impacted Louisiana Cases, Nov 2013 Sub (1973–2011varying) - Linked To County Attributes - Total U.S., 1969–2012 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2014 (updated 5/7/2014), based on the November 2013 submission.
3. Independent U.K. Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380: 1778–86. [PubMed: 23117178]
4. Nelson HD, Tyne K, Naik A, Bougatsos BS, Chan BK, Humphrey L. Screening for breast cancer: An update for the U.S. Preventive Services Task Force. Ann Intern Med 2009;151:727–37. [PubMed: 19920273]
5. Tabar L, Vitak B, Chen TH, Yen AM, Cohen A, Tot T, et al. Swedish two-country trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011;260:658–63. [PubMed: 21712474]
6. Paci E, Euroscreen Working Group. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. J Med Screen 2012;19Suppl 1:5–13. [PubMed: 22972806]
7. International Agency for Research on Cancer. European code against cancer. Lyon, France: International Agency for Research on Cancer; 2014.
8. Taplin SH, Ichikawa L, Yood MU, Manos MM, Geiger AM, Weinmann S, et al. Reason for late-stage breast cancer: absence of screening or detection, or breakdown in follow-up? J Natl Cancer Inst 2004;96:1518–27. [PubMed: 15494602]
9. Malmgren JA, Parikh J, Atwood MK, Kaplan HG. Impact of mammography detection on the course of breast cancer in women aged 40–49. Radiology 2012;262:797–806. [PubMed: 22357883]
10. Nickson C, Mason KE, English DR, Kavanagh AM. Mammographic screening and breast cancer mortality: A case-control study and meta-analysis. Cancer Epidemiol Biomarkers Prev 2012;21:1479–88. [PubMed: 22956730]
11. Final Update Summary: Breast Cancer: Screening. U.S. Preventive Services Task Force. 1 2016 [cited 2016 Feb. 14]. Available at http://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/breast-cancer-screening.
12. Smith RA, Brooks D, Cokkinides V, Salsow D, Brawley OW. Cancer screening in the United States, 2013: a review of current American Cancer Society guidelines, current issues in cancer screening, and new guidance on cervical cancer screening and lung cancer screening. CA Cancer J Clin 2013;63:87–105.
13. Fenton JJ, Barton MB, Geiger AM, Herrinton LJ, Rolnick SJ, Harris EL, et al. Screening clinical breast examination: how often does it miss lethal breast cancer? J Natl Cancer Inst Monogr 2005;35:67–71.
14. Mandelson MT, Oestreicher N, Porter JL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 2000;92: 1081–7. [PubMed: 10880551]
15. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 1996;276:33–8. [PubMed: 8667536]
16. Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late-stage diagnosis of breast and cervical cancer. Am J Public Health 1991;81:646–9. [PubMed: 2014871]

17. Hunter CP, Redmond CK, Chen VW, Austin DF, Greenberg RS, Correa P, et al. Breast cancer: factors associated with stage at diagnosis in black and white women. J Natl Cancer Inst 1993;85:1129–37. [PubMed: 8320742]
18. Virnig BA, Baxter NN, Habermann EB, Feldman RD, Bradley CJ. A matter of race: early- versus late-stage cancer diagnosis. Health Affairs 2009;28:160–8. [PubMed: 19124866]
19. Moorman PG, Jones BA, Millikan RC, Hall IJ, Newman B. Race, anthropometric factors, and stage at diagnosis of breast cancer. Am J Epidemiol 2001;153:284–91. [PubMed: 11157416]
20. Roetzheim RG, Pal N, Tennant C, Voti L, Ayanian JZ, Schwabe A, et al. Effects of health insurance and race on early detection in cancer. J Natl Cancer Inst 1999;91:1409–15. [PubMed: 10451447]
21. Halpern MT, Bian J, Ward EM, Schrag NM, Chen AY. Insurance status and stage of cancer diagnosis among women with breast cancer. Cancer 2007; 110:403–11. [PubMed: 17562557]
22. Smith-Bindman R, Miglioretti DL, Lurie N, Abraham L, Ballard-Barbash R, Strzelczyk J, et al. Does utilization of screening mammography explain racial and ethnic differences in breast cancer? Ann Intern Med 2006; 144:541–53. [PubMed: 16618951]
23. Lantz PM, Mujahid M, Schwartz K, Janz NK, Fagerlin A, Salem B, et al. The influence of race, ethnicity, and individual socioeconomic factors on breast cancer stage at diagnosis. Am J Public Health 2006;96:2173–8. [PubMed: 17077391]
24. McCarthy EP, Burns R, Coughlin SS, Freund KM, Rice J, Marwill SL. Mammography use helps to explain differences in breast cancer stage at diagnosis between older black and white women. Ann Intern Med 1998;128:729–36. [PubMed: 9556466]
25. Sassi F, Luft HS, Guadagnoli E. Reducing racial/ethnic disparities in female breast cancer: screening rates and stage at diagnosis. Am J Public Health 2006;96:2165–72. [PubMed: 17077392]
26. Hahn KME, Bondy ML, Selvan M, Lund MJ, Liff JM, Flagg EW, et al. Factors associated with advanced disease stage at diagnosis in a population-based study of patients with newly diagnosed breast cancer. Am J Epidemiol 2007;166:1035–44. [PubMed: 17690220]
27. Lannin DR, Mathews HF, Mitchell J, Swanson MS, Swanson FH, Edwards MS. Influence of socioeconomic and cultural factors on racial differences in late-stage presentation of breast cancer. JAMA 1998;279:1801–7. [PubMed: 9628711]
28. Bantina NG, Trentham-Dietz A, Gangnon RE, Sprague BL, Rosenberg MA, Stout NK, et al. Variation in tumor natural history contributes to racial disparities in breast cancer stage at diagnosis. Breast Cancer Res Treat 2013;138:519–28. [PubMed: 23417335]
29. Jones BA, Kasl SV, Curnen MGM, Owens PH, Dubrow R. Can mammography screening explain the race difference in stage at diagnosis? Cancer 1995;75:2103–13. [PubMed: 7697601]
30. Wells KJ, Roetzheim RG. Health disparities in receipt of screening mammography in Latinas: a critical review of recent literature. Cancer Control 2007;14:369–79. [PubMed: 17914337]
31. U.S. Centers for Disease Control and Prevention. Cancer Screening – United States, 2010. Morb Mortal Wkly Rep (MMWR) 2012;61:41–5. [PubMed: 22278157]
32. Sabatino SA, Coats RJ, Uhler RJ, Breen N, Tangka F, Shaw KM. Disparities in mammography use among US women aged 40–64 years, by race, ethnicity, income, and health insurance status. Med Care 2008;46:692–700. [PubMed: 18580388]
33. Ward E, Halpern M, Schrag N, Cokkinides V, DeSantis C, Bandi P, et al. Association of insurance with cancer care utilization and outcomes. CA Cancer J Clin 2008;58:9–31. [PubMed: 18096863]
34. Bradley CJ, Given CW, Roberts C. Late stage cancers in a Medicaid-insured population. Med Care 2003;41:722–8. [PubMed: 12773838]
35. Ayanian JZ, Kohler BA, Abe T, Epstein AM. The relation between health insurance coverage and clinical outcomes among women with breast cancer. N Engl J Med 1993;329:326–31. [PubMed: 8321261]
36. Fleming ST, Pursley HG, Newman B, Pavlov D, Chen K. Comorbidity as a predictor of stage of illness for patients with breast cancer. Med Care 2005;43:132–40. [PubMed: 15655426]
37. Yasmeen S, Rubbard R, Romano PS, Zhu W, Geller BM, Onega T, et al. Risk of advanced-stage breast cancer among older women with comorbidities. Cancer Epidemiol Biomarkers Prev 2012;21:1510–9. [PubMed: 22744339]

38. Schairer C, Li Y, Frawley P, Graubard BI, Wellman RD, Buist DSM, et al. Risk factors for inflammatory breast cancer and other invasive breast cancers. J Natl Cancer Inst 2013;105:1373–84. [PubMed: 24046390]
39. Kerlikowske K, Cook AJ, Buist DSM, Cummings SR, Vachon C, Vacek P, et al. Breast cancer risk by breast density, menopause, and postmenopausal hormone therapy use. J Clin Oncol 2010;28:3830–7. [PubMed: 20644098]
40. Taplin SH, Yabroff KR, Zapka J. A multilevel research perspective on cancer care delivery: the example of follow-up to an abnormal mammogram. Cancer Epidemiol Biomarkers Prev 2012;21:1709–15. [PubMed: 22911332]
41. Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. JAMA 2001;286:2143–51. [PubMed: 11694156]
42. Pizer SD. An intuitive review of methods for observational studies of comparative effectiveness. Health Serv Outcomes Res Method 2009;9: 54–68.
43. Heckman J. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. J Hum Resour 1997;32: 441–62.
44. Rassen JA, Brookhard MA, Gynn RJ, Mittleman Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. J Clin Epidemiol 2009;62:1226–32. [PubMed: 19356901]
45. Shih YT, Elting LS, Halpern MT. Factors associated with immunotherapy use among newly diagnosed cancer patients. Med Care 2009;47:948–58. [PubMed: 19704352]
46. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. J Health Econ 2008;27:531–43. [PubMed: 18192044]
47. Hadley J, Yabroff KR, Barrett MJ, Penson DF, Saigal CS, Potosky AL. Comparative effectiveness of prostate cancer treatments: evaluating statistical adjustments for confounding in observational data. J Natl Cancer Inst 2010;102:1780–93. [PubMed: 20944078]
48. German RR, Wike JM, Bauer KR, Fleming ST, Trentham-Dietz A, Namiak M, et al. Quality of cancer registry data: findings from the CDC-NPCR's Breast and Prostate Cancer Data Quality and Patterns of Care Study. J Registry Manag 2011;38:75–86. [PubMed: 22096878]
49. Piccirillo JF, Creech C, Zequeira R, Anderson S, Johnson AS. Inclusion of comorbidity into oncology data registries. J Registry Manag 1999;26:66–70.
50. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295:2492–502 [PubMed: 16757721]
51. Garrido MM, Deb P, Burgess JF Jr, Penrod JD. Choosing models for health care cost analyses: issues of nonlinearity and endogeneity. HSR: Health Services Research 2012;47:2377–97. [PubMed: 22524165]
52. American Cancer Society. Breast cancer facts & figures 2013–2014. Atlanta: American Cancer Society, Inc; 2013.
53. Staiger D, Stock J. Instrumental variables regression with weak instruments. Econometrica 1997;65:557–86.
54. Allison PD. Missing data. Thousand Oaks, CA: Sage; 2001.
55. StataCorp. Stata Release 13. Statistical Software. College Station, TX: StataCorp LP; 2013.
56. Winship C, Radbill L. Sampling weights and regression analysis. Sociol Methods Res 1994;23:230–57.
57. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 2000.
58. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757–63. [PubMed: 9382394]
59. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of treatment to non-randomized control group (Tutorial in Biostatistics). Statist Med 1998;17:2265–81.
60. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. Am J Epidemiol 2006;163:262–70. [PubMed: 16371515]

61. Brooks JM, Ohsfeldt RL. Squeezing the balloon: propensity scores and unmeasured covariate balance. HSR: Health Services Research 2013;48: 1487–507. [PubMed: 23216471]
62. Wright JD, Ananth CV, Tsui J, Glied SA, Burke WM, Lu Y-S, et al. Comparative effectiveness of upfront treatment strategies in elderly women with ovarian cancer. Cancer 2014;120:1246–54. [PubMed: 24443159]
63. Shen Y, Yang Y, Inoue LYT, Munsell MF, Miller AB, Berry DA. Role of detection method in predicting breast cancer survival: analysis of randomized screening trials. J Natl Cancer Inst 2005;97:1195–203. [PubMed: 16106024]
64. Reeves MJ, Newcomb PA, Remington PL, Marcus PM, MacKenzie WR. Body mass and breast cancer: relationship between method of detection and stage of disease. Cancer 1996;77:301–7. [PubMed: 8625238]
65. Riley GF, Potosky AL, Klabunde CN, Warren JL, Ballard-Barbash R. Stage at diagnosis and treatment patterns among older women with breast cancer: an HMO and fee-for-service comparison. JAMA 1999;281:720–6. [PubMed: 10052442]
66. Lee-Feldstein A, Feldstein PJ, Buchmueller T, Katterhagen G. The relationship of HMOs, health insurance, and delivery systems to breast cancer outcomes. Med Care 2000;38:705–18. [PubMed: 10901354]
67. Elmore JG, Nakano CY, Linden HM, Reisch LM, Ayanian JZ, Larson EB. Racial inequities in the timing of breast cancer detection, diagnosis, and initiation of treatment. Med Care 2005;43:141–8. [PubMed: 15655427]
68. Clegg LX, Reichman ME, Miller BA, Hankey BF, Singh GK, Lin YD, et al. Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the Surveillance, Epidemiology and End Results: National Longitudinal Mortality Study. Cancer Causes Control 2009;20: 417–35. [PubMed: 19002764]
69. Byers TE, Wolf HJ, Bauer KR, Bolick-Aldrich S, Chen VW, Finch JL, et al. The impact of socioeconomic status on survival after cancer in the United States. Cancer 2008;113:582–91. [PubMed: 18613122]
70. Barry J, Breen N, Barrett M. Significance of increasing poverty levels for determining late-stage breast diagnosis in 1990 and 2000. J Urban Health 2012;89:614–27. [PubMed: 22322332]
71. Kerlikowske K, Walker R, Miglioretti DL, Desai A, Ballard-Barbash R, Buist DSM; for the National Cancer Institute-Sponsored Breast Cancer Surveillance Consortium. Obesity, mammography use and accuracy, and advanced breast cancer risk. J Natl Cancer Inst 2008;100:1724–33. [PubMed: 19033562]
72. Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D, et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. JAMA Intern Med 2013; 173:807–16. [PubMed: 23552817]
73. Maruthur NM, Bolen S, Brancati FL, Clark JM. Obesity and mammography: a systematic review and meta-analysis. J Intern Med 2009;24:665–77.
74. Gapstur SM, Lopez P, Colangelo LA, Wolfman J, Van Horn L, Hendrick RE. Associations of breast cancer risk factors with breast cancer in Hispanic women. Cancer Epidemiol Biomarkers Prev 2003;12:1074–80. [PubMed: 14578145]
75. Green LE, Dinh TA, Smith RA. An estrogen model: the relationship between body mass index, menopausal status, estrogen replacement therapy, and breast cancer risk. Comput Math Methods Med 2012;12, Article ID 792375, 8pages.
76. Cui Y, Whiteman MK, Flaws JA, Langenberg P, Tkaczuk KH, Bush TL. Body mass and stage of breast cancer diagnosis. Int J Cancer 2002;98: 279–83. [PubMed: 11857420]
77. Hunt KA, Sickles EA. Effect of obesity on screening mammography: outcomes analysis of 88,346 consecutive examinations. AJR 2000;174: 1251–5. [PubMed: 10789771]

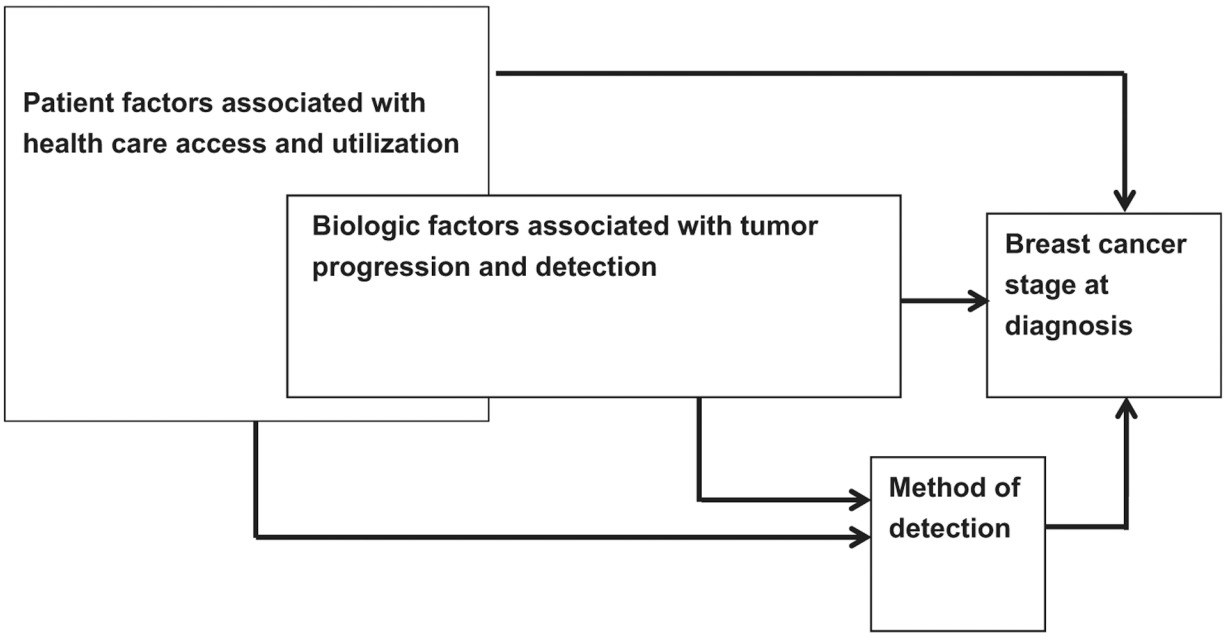

Figure 1.
Breast cancer stage at diagnosis: conceptual framework.

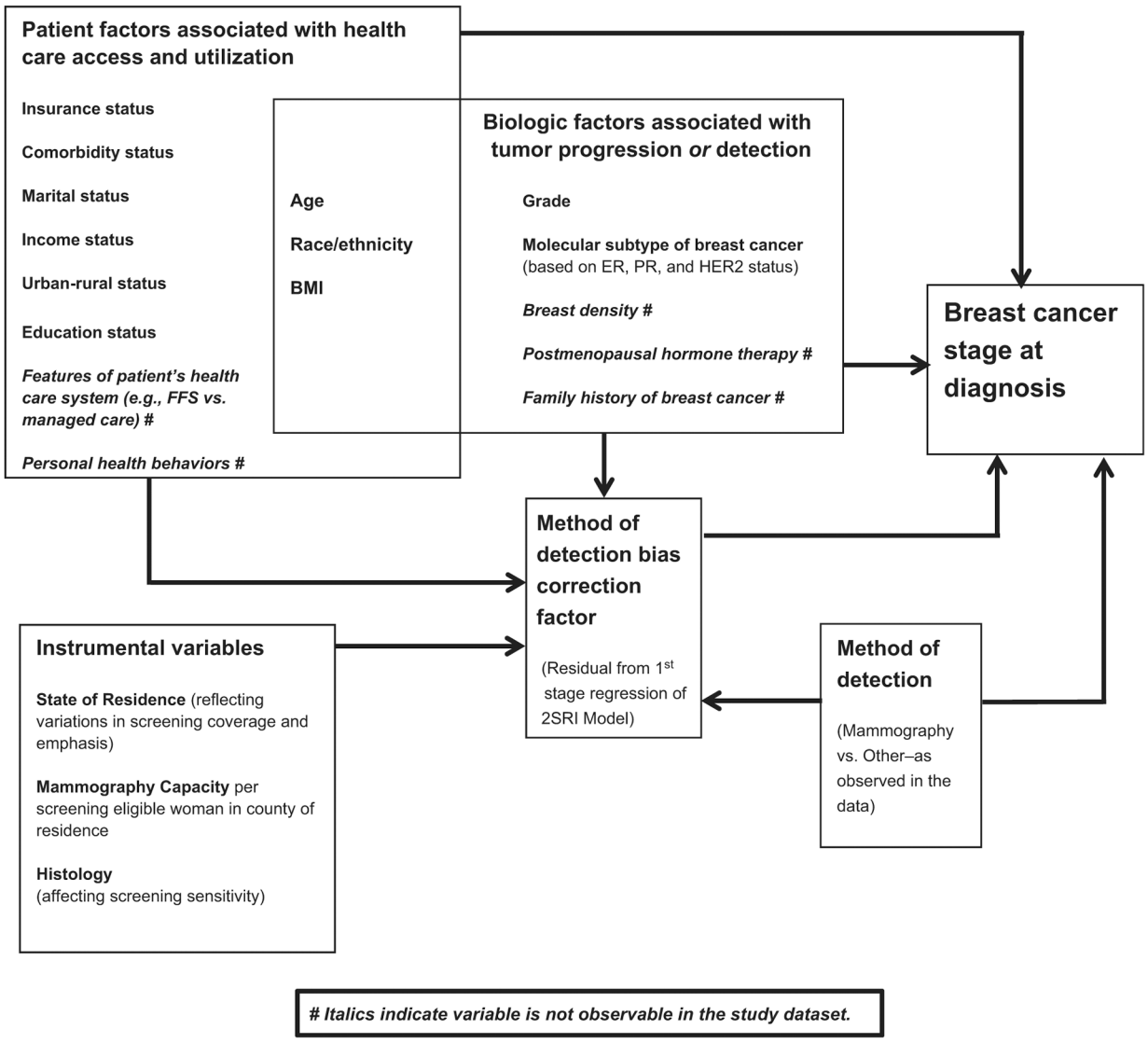

Figure 2.

Conceptualizing the statistical analysis of the determinants of stage: transforming the framework to reflect estimation via 2SRI IV model.

Table 1.

Percent distribution of patients across predictor variable levels and percent distribution by stage at diagnosis (advanced vs. earlier[a]) within each variable level, along with unadjusted ORs for advanced stage

Variable	Total	Advanced stage (*N* = 762)	Earlier stage (*N* = 6,741)	OR (advanced vs. earlier stage)	*P*	95% CI
Method of detection [b]						
Mammography	3,718 (49.6%)	2.6%	97.4%	0.13	<0.001	0.10–0.16
Other	3,785 (50.4%)	17.6%	82.4%	Ref		
Age group						
<40	497 (6.6%)	13.1%	86.9%	1.37	0.03	1.03–1.83
40–49	1,681 (22.4%)	9.1%	90.9%	0.91	0.39	0.74–1.12
50–64	2,823 (37.6%)	9.9%	90.1%	Ref		
65–69	740 (9.9%)	8.8%	91.2%	0.88	0.37	0.66–1.17
70–79	1,187 (15.8%)	10.4%	89.6%	1.05	0.65	0.84–1.32
≥80	575 (7.7%)	13.4%	86.6%	1.41	0.01	1.08–1.85
Mole subtype [c]						
Luminal A	4,671 (62.3%)	10.2%	89.8%	Ref		
Triple negative	1,307 (17.4%)	14.4%	85.6%	1.92	<0.001	1.58–2.34
HER2 type	951 (12.7%)	10.6%	89.4%	1.55	0.002	1.18–2.03
Luminal B	574 (7.7%)	17.4%	82.6%	2.23	<0.001	1.70–2.92
Grade [c]						
Well differentiated	1,365 (18.2%)	2.8%	97.2%	Ref.		
Moderately differentiated	2,972 (39.6%)	7.6%	92.4%	2.84	<0.001	1.87–4.32
Poorly or undifferentiated	3,166 (42.2%)	15.7%	84.3%	6.50	<0.001	4.33–9.77
Comorbidity status						
None	3,214 (42.8%)	8.8%	91.2%	Ref		
Mild	3,270 (43.6%)	9.9%	90.1%	1.14	0.12	0.97–1.35
Moderate	723 (9.6%)	13.4%	86.6%	1.61	<0.001	1.26–2.05
Severe	296 (3.9%)	19.3%	80.7%	2.47	<0.001	1.81–3.38
Race/ethnicity						
White	4,009 (53.4%)	8.2%	91.8%	Ref.		

Variable	Total	Advanced stage (*N* = 762)	Earlier stage (*N* = 6,741)	OR (advanced vs. earlier stage)	*P*	95% CI
Black	2,309 (30.8%)	13.9%	86.1%	1.81	<0.001	1.53–2.13
Asian/Pacific Islander (API)	438 (5.8%)	9.1%	90.9%	1.13	0.49	0.80–1.59
American Indian/Alaska Native (AI/AN)	55 (0.7%)	10.9%	89.1%	1.38	0.47	0.58–3.23
Hispanic	692 (9.2%)	9.8%	90.2%	1.23	0.15	0.93–1.61
Education status [d]						
<25% w/o HS educ	4,314 (57.5%)	8.4%	91.6%	Ref.		
≥25% w/o HS educ	3,189 (42.5%)	12.5%	87.5%	1.55	<0.001	1.34–1.80
Marital status						
Married	4,209 (56.1%)	8.2%	91.8%	Ref.		
Single	1,141 (15.2%)	14.4%	85.6%	1.87	<0.001	1.53–2.28
Separated/widowed/divorced	2,153 (28.7%)	11.6%	88.4%	1.45	<0.001	1.22–1.72
Poverty status [d]						
<20% below federal level	5,510 (73.4%)	8.8%	91.2%	Ref.		
≥20% below federal level	1,993 (26.6%)	13.7%	86.3%	1.64	<0.001	1.40–1.92
Urban/rural status [d]						
Urban	3,841 (51.2%)	10.5%	89.5%	Ref.		
Rural	1,008 (13.4%)	10.9%	89.1%	1.06	0.70	0.84–1.31
Urban-rural mix	2,654 (35.4%)	9.4%	90.6%	0.88	0.13	0.74–1.04
Insurance						
Private	4,542 (60.5%)	7.6%	92.4%	Ref.		
Uninsured	231 (3.1%)	19.9%	80.1%	2.96	<0.001	2.10–4.16
Medicaid	1,089 (14.5%)	17.0%	83.0%	2.43	<0.001	2.01–2.95
Medicare	1,641 (21.9%)	10.9%	89.1%	1.46	<0.001	1.20–1.76
BMI [c]						
<25	2,348 (31.3%)	9.8%	90.2%	Ref.		
25–29.9	2,158 (28.8%)	8.9%	91.1%	0.98	0.85	0.78–1.23
30–39.9	2,444 (32.6%)	10.2%	89.8%	1.13	0.27	0.91–1.41
≥40	553 (7.4%)	16.3%	83.7%	1.91	<0.001	1.42–2.57

[a]Advanced stage includes AJCC stages IIIB, IIIC, and IV; earlier stage includes AJCC stages 0, I, IIA, IIB, and IIIA.

[b] Data abstractors were instructed to search the patient's medical records to identify the method by which the breast cancer was "initially detected," which refers to the first notice of the cancer, not the diagnostic procedures that may have followed. Detection by mammography meant a "routine" screen for women without a personal history of breast cancer. CBE was assigned if the breast cancer was detected through a routine clinical breast exam by a health care provider. BSE was assigned if the cancer was detected by a routine breast self-exam or an exam by a partner. The assigned method was signs/symptoms when the patient was recorded as having nipple discharge, bleeding, dimpling, or other abnormal signs/symptoms (other than a lump); also included in this variable category were cancers detected incidental to the discovery of a breast mass on a workup for another condition, and instances in which no recorded method could be found in the records. In some cases, assigning the method of detection depends on evaluating the sequence and clinical connectedness of events. For example, if a woman has a routine screening mammogram and, upon the next visit to the physician, learns that it is positive and that the pain she has begun to feel in her breast could be cancerous, the method assigned would be mammography—the "event" that triggered the flow of data that followed. On the other hand, if the record accompanying a mammogram noted that a precipitator of the screen was breast-related pain or symptoms, the encoded method would be signs/symptoms (even if the mammogram then performed is positive). POC-BP data abstractors underwent extensive didactic and experiential training prior to the launch of data collection.

[c] Missing values for mole-subtype, grade, and BMI were imputed (see text) and included in this base-case sample. No imputation was performed for the other predictor variables, which collectively were missing very few observations; but any patient missing a value on these (nonimputed) variables was excluded from the base-case sample, which ultimately included 7,503 patients with no missing predictor variable values.

[d] An area-level variable constructed from 2000 US Census data by linking each patient to a census tract based on her residential address; the patient was then assigned the value (level) of the variable applicable to her census tract.

Table 2.

Percent distribution of patients across instrument variable levels and percent distribution by method of detection (mammography vs. other methods) within each variable level, along with unadjusted ORs for detection by mammography

IV	Total	Mammography (*N* = 3,718)	Other methods (*N* = 3,785)	OR (mammography vs. other methods)	*P*	95% CI
State						
LA	1,622 (21.6%)	48.6%	51.4%	Ref.		
WI	974 (13.0%)	57.9%	42.1%	1.46	<0.001	1.24–1.71
NC	964 (12.8%)	45.9%	54.1%	0.90	0.18	0.76–1.05
CA	1,499 (20.0%)	46.8%	53.2%	0.93	0.31	0.81–1.07
KY	438 (5.8%)	47.5%	52.5%	0.96	0.69	0.78–1.18
GA	2,006 (26.7%)	50.6%	49.4%	1.08	0.23	0.95–1.24
Histology [a]						
Lobular	427 (5.7%)	39.6%	60.4%	Ref.		
Ductal or ductal/lobular	5,462 (72.8%)	48.3%	51.7%	1.43	<0.001	1.17–1.74
Other	1,614 (21.5%)	56.5%	43.5%	1.98	<0.001	1.60–2.47
Mammography-capacity [b]	7,503	1.308	1.248	1.17	<0.001	1.09–1.26

[a] Derived from an underlying, POC-BP generated 10-level categorization of histology codes, in which lobular breast cancers were distinguished from a combined category consisting of both ductal and ductal/lobular tumors and from a composite “other” category consisting of mucinous, tubular, comedocarcinoma, inflammatory, medullary, papillary, and all other tumors.

[b] A county-level, continuous variable defined as the estimated annual mammographic-capacity of the county (number of machines X an assumed capability of 6,000 mammographic exams/year) divided by the number of women aged over 40 in the county. Shown here are the means of mammography-capacity by method of detection. The underlying source of county-level data on mammographic-capacity was the federally supported Mammography Program Reporting and Information System (MPRIS), which provided data for a collaborative effort by the FDA and the CDC, initiated in 2008, to document facilities conducting mammography (and their machine capacity) at the US county level. Among analyses using the MPRIS was one conducted by the U.S. Government Accountability Office (U.S. GAO Mammography: Current nationwide capacity is adequate, but access problems may exist in certain locations. Washington, D.C: U.S. GAO, July 2006), in which it was assumed that one machine and one radiologic technologist could perform three mammograms per hour; from that was derived the assumption of 6,000 mammograms per machine per year. The MPRIS data for our estimates included the 2003–2005 period. As noted, the mammographic-capacity variable could not be created for patients residing in Minnesota, where state legislation and accompanying regulations continue to “prohibit the release of count-level data to outside entities” (http://www.statecancerprofiles.cancer.gov/datanotavailable.html). The Minnesota state cancer registry determined that the CDC constituted such an “outside entity,” notwithstanding the state’s commitment to participate in the CDC-funded POC-BP study.

Table 3.

Determinants of an advanced-stage diagnosis of breast cancer: single-equation and 2SRI base-case models (N = 7,503)

	Single-equation model				2SRI model			
	c = 0.796	P value of H-L test = 0.53			c = 0.797	P value of H-L test = 0.88		
Predictor variables	**OR**	***P***	**95% CI**		**OR**	***P***	**95% CI**	
Method of detection								
Other	Ref.				Ref.			
Mammography	0.14	<0.001	0.11	0.18	0.04	<0.001	0.01	0.11
First-stage residual	N/A				3.89	0.02	1.27	11.90
Age group								
50–64	Ref.				Ref.			
<40	0.72	0.04	0.52	0.98	0.48	0.001	0.32	0.73
40–49	0.67	<0.001	0.53	0.84	0.57	<0.001	0.42	0.78
65–69	1.04	0.83	0.75	1.43	1.13	0.53	0.77	1.65
70–79	1.15	0.34	0.87	1.51	1.16	0.33	0.86	1.58
>80	1.24	0.22	0.88	1.73	1.10	0.56	0.79	1.54
Mole-subtype								
Luminal A	Ref.				Ref.			
Triple negative	0.94	0.61	0.75	1.18	0.84	0.18	0.65	1.09
HER2 Type	1.34	0.04	1.01	1.77	1.40	<0.001	1.17	1.67
Luminal B	1.32	0.07	0.98	1.79	1.28	0.09	0.96	1.69
Grade								
Well differentiated	Ref.				Ref.			
Moderately differentiated	2.48	<0.001	1.63	3.79	2.36	<0.001	1.70	3.28
Poorly or undifferentiated	4.64	<0.001	3.01	7.16	3.91	<0.001	2.77	5.53
Comorbidity status								
None	Ref.				Ref.			
Mild	1.05	0.66	0.86	1.27	1.08	0.47	0.88	1.32
Moderate	1.11	0.50	0.83	1.49	1.08	0.66	0.78	1.48
Severe	1.46	0.04	1.02	2.10	1.29	0.21	0.87	1.90
Race/ethnicity								

	Single-equation model				2SRI model			
	c = 0.796	*P* value of H-L test = 0.53			*c* = 0.797	*P* value of H-L test = 0.88		
Predictor variables	**OR**	***P***	**95% CI**		**OR**	***P***	**95% CI**	
White	Ref.				Ref.			
Black	1.19	0.10	0.97	1.46	1.16	0.07	0.99	1.36
API	1.01	0.96	0.69	1.48	0.94	0.76	0.63	1.40
AI/AN	1.25	0.61	0.53	2.97	1.02	0.98	0.36	2.91
Hispanic	0.97	0.85	0.72	1.32	0.89	0.40	0.67	1.17
Education status								
<25% w/o HS educ	Ref.				Ref.			
>25% w/o HS educ	1.09	0.40	0.89	1.34	1.06	0.56	0.88	1.26
Marital status								
Married	Ref.				Ref.			
Single	1.34	0.01	1.07	1.69	1.28	0.06	0.99	1.65
Separated/widowed/divorced	1.09	0.38	0.89	1.34	1.03	0.79	0.81	1.33
Poverty level								
<20% below federal level	Ref.				Ref.			
>20% below federal level	1.12	0.32	0.90	1.39	1.13	0.24	0.92	1.38
Urban/rural status								
Urban	Ref.				Ref.			
Rural	1.06	0.68	0.81	1.37	1.03	0.82	0.80	1.33
Urban-rural mix	0.96	0.70	0.80	1.16	0.98	0.85	0.82	1.18
Insurance								
Private	Ref.				Ref.			
Uninsured	1.85	0.001	1.28	2.69	1.60	0.07	0.96	2.66
Medicaid	1.63	<0.001	1.31	2.04	1.48	0.01	1.09	2.02
Medicare	1.33	0.02	1.04	1.69	1.29	0.04	1.02	1.65
BMI								
<25	Ref.				Ref.			
25–29.9	0.99	0.94	0.78	1.27	1.04	0.68	0.85	1.28
30–39.9	1.13	0.34	0.88	1.45	1.20	0.08	0.98	1.47
>40	1.52	0.02	1.07	2.17	1.62	0.001	1.23	2.13

Table 4.

Race/ethnicity, insurance, and the relative odds of an advanced-stage diagnosis of breast cancer, across prediction models[a]

	Race/ethnicity (white is reference category)				Insurance (private is reference category)		
	Black	**Hispanic**	**API**	**AI/AN**	**Uninsured**	**Medicaid**	**Medicare**
#1. Bivariate logistic (race/ethnicity)	1.81 (<0.001)	1.23 (0.15)	1.13 (0.49)	1.38 (0.47)			
#2. Bivariate logistic (insurance)					2.96 (<0.001)	2.43 (<0.001)	1.46 (<0.001)
#3. Two-variable logistic (race/ethnicity and insurance)	1.56 (<0.001)	1.14 (0.37)	1.09 (0.64)	1.11 (0.82)	2.71 (<0.001)	2.16 (<0.001)	1.46 (<0.001)
#4. Single-equation w/o bio variables (Supplementary Table S1)	1.31 (0.01)	0.97 (0.83)	0.98 (0.99)	1.14 (0.76)	2.06 (<0.001)	1.65 (<0.001)	1.32 (0.02)
#5. Single-equation with bio variables (Table 3)	1.19 (0.01)	0.97 (0.85)	1.01 (0.96)	1.25 (0.61)	1.85 (0.001)	1.63 (<0.001)	1.33 (0.02)
#6. 2SRI w/o bio variables (Supplementary Table S1)	1.26 (0.03)	0.89 (0.52)	0.92 (0.72)	0.95 (0.93)	1.73 (0.06)	1.49 (0.02)	1.29 (0.06)
#7. 2SRI with bio variables (base-case model; Table 3)	1.16 (0.07)	0.89 (0.40)	0.94 (0.76)	1.02 (0.98)	1.60 (0.07)	1.48 (0.01)	1.29 (0.04)
Additional sensitivity analyses							
#8. 2SRI base-case w/o missing value imputation (Supplementary Table S2)	1.28 (0.10)	1.00 (0.99)	1.14 (0.64)	0.79 (0.71)	1.62 (0.10)	1.30 (0.12)	1.27 (0.22)
#9. 2SRI base-case with sample weights (Supplementary Table S3)	1.23 (0.10)	1.04 (0.84)	1.01 (0.96)	0.96 (0.94)	1.87 (0.02)	1.35 (0.11)	1.22 (0.24)
#10. 2SRI model excluding mammography-capacity (Supplementary Table S4)	1.16 (0.09)	0.89 (0.44)	0.96 (0.81)	1.44 (0.36)	1.72 (0.003)	1.57 (<0.001)	1.36 (0.02)
#11. Propensity score-adjusted model via IPTW (Supplementary Table S5)	1.22 (0.07)	1.05 (0.79)	0.94 (0.77)	1.09 (0.84)	1.93 (0.002)	1.68 (<0.001)	1.38 (0.02)
#12. Propensity score-adjusted model via SMRW (Supplementary Table S5)	1.20 (0.12)	1.001 (0.99)	0.88 (0.55)	0.98 (0.97)	2.06 (0.002)	1.66 (<0.001)	1.38 (0.02)

[a]Table entries are logistic model-derived ORs, with corresponding *P* values in parentheses. ORs in rows **#1–3** are derived from logistic models with either one (race/ethnicity or insurance) or two predictor variables (both race/ethnicity and insurance) estimated with the 7,503 observations used in the base-case multivariable models from Table 3; ORs in rows **#4–12** are taken directly from the indicated tables.